JACS Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

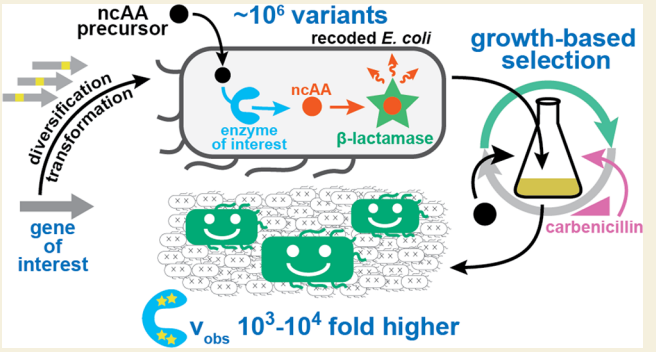
pubs.acs.org/jacsau
Article

A Robust Growth-Based Selection Platform to Evolve an Enzyme via Dependency on Noncanonical Tyrosine Analogues

Suzanne C. Jansen and Clemens Mayer*

Cite This: JACS Au 2024, 4, 1583−1590
Read Online

ABSTRACT: Growth-based selections evaluate the fitness of individual organisms at a population level. In enzyme engineering, such growth selections allow for the rapid and straightforward identification of highly efficient biocatalysts from extensive libraries. However, selection-based improvement of (synthetically useful) biocatalysts is challenging, as they require highly dependable strategies that artificially link their activities to host survival. Here, we showcase a robust and scalable growth-based selection platform centered around the complementation of noncanonical amino acid-dependent bacteria. Specifically, we demonstrate how serial passaging of populations featuring millions of carbamoylase variants autonomously selects biocatalysts with up to 90,000-fold higher initial rates. Notably, selection of replicate populations enriched diverse biocatalysts, which feature distinct amino acid motifs that drastically boost carbamoylase activity. As beneficial substitutions also originated from unintended copying errors during library preparation or cell division, we anticipate that our growth-based selection platform will be applicable to the continuous, autonomous evolution of diverse biocatalysts in the future.



KEYWORDS: *enzyme engineering, directed evolution, growth-based selections, high throughput, carbamoylases, biocatalysis, synthetic biology*

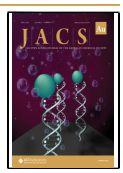
INTRODUCTION

Evolution is a dynamic, all-purpose problem solver that enzyme engineers mimic in the laboratory to generate tailor-made biocatalysts.[1−4] In general, such directed evolution campaigns rely on four principal steps: (1) generating genetic diversity, (2) translating the genetic information into biocatalysts, (3) identifying variants with desired properties, and (4) recovering genes encoding improved enzymes (Figure 1A).[5−7] While concurrent molecular biology tools readily enable the creation, translation, and recovery of vast gene libraries, exhaustive sampling of the resulting sequence space remains a persistent bottleneck for most directed evolution campaigns.[8,9] Therefore, general strategies to identify improved biocatalysts are highly sought-after. Ideally, such strategies should not only be fast, accurate, cheap, and easy to implement, but also allow for a high throughput and have the ability to differentiate enzymes with vastly different activities (dynamic range, Figure 1B).

In principle, growth-based in vivo selections—also referred to as growth-coupled or life-or-death selections—have the potential to meet all these criteria.[10−12] First, they can rapidly assess all library members of vast populations simultaneously by applying a selective pressure that effectively eliminates undesired enzyme variants (Figure 1C). Moreover, growth-based selections are cheap and straightforward to implement in the laboratory, as identifying improved biocatalysts typically involves either plating populations on selective media or serially (or continuously) culturing populations under selective conditions. In the latter scenario, more active biocatalysts will gradually outcompete less proficient ones (Figure 1D).[13−16] Lastly, previously employed selections for biocatalysts that directly confer a growth advantage to a host organism (e.g., chorismate mutase or β-lactamase) could be employed to identify enzyme variants with a wide range of activities.[17−19]

Despite these advantages, growth-based selections are rarely employed for the directed evolution of synthetically useful enzymes that function on small molecules, as such biocatalysts typically do not confer a growth advantage to a producing organism. Instead, the use of (synthetic) auxotrophs,[20,21] genetic circuits,[22,23] or other sensors[24,25] is necessary to artificially link the activity of a desired biocatalyst to cell fitness/survival. While such *chemical complementation systems* are effective in interrogating individual enzyme variants of

Received: January 23, 2024
Revised: February 21, 2024
Accepted: February 22, 2024
Published: March 19, 2024



https://doi.org/10.1021/jacsau.4c00070
JACS Au 2024, 4, 1583−1590

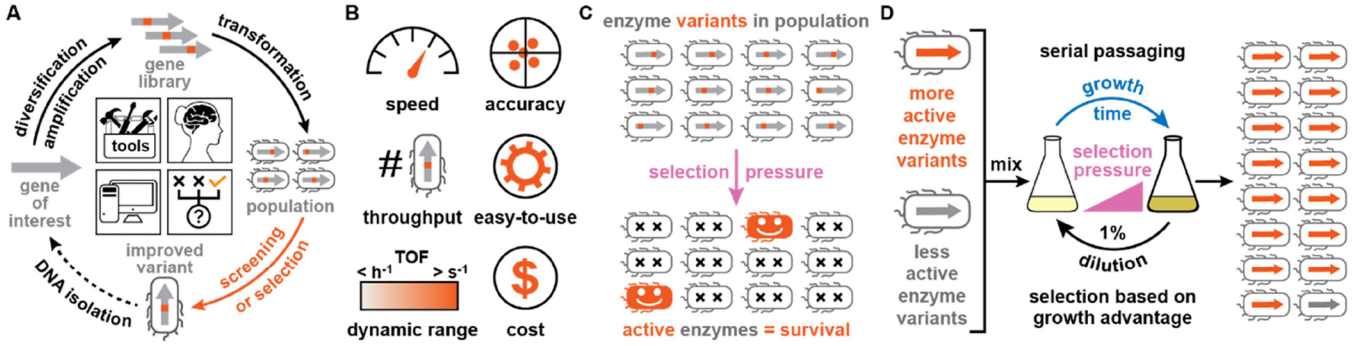


Figure 1. Directed evolution of biocatalysts by growth-based selections. (A) To tailor a gene of interest, directed evolution approaches rely on iterative cycles of genetic diversification, translation, and identification and isolation of improved enzyme variants. (B) Strategies to identify improved biocatalysts should ideally meet a broad set of criteria. (C) By application of a tailored selection pressure, selections evaluate all enzyme variants in a population simultaneously and remove undesired variants. (D) Serial passaging under selection pressure can amplify more active enzyme variants in a mixed population by virtue of the growth advantage they bestow on their host. Panel C was reproduced with permission from ref 16. Copyright 2022, Wiley-VCH GmbH.

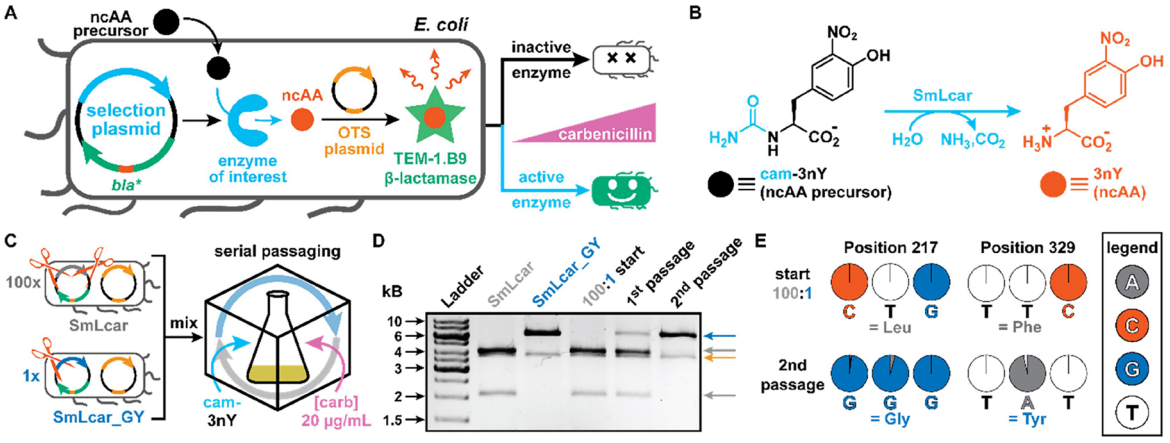


Figure 2. Selecting improved variants from heavily skewed populations. (A) Scheme of our growth-based selection platform that makes use of chemical complementation. An enzyme of interest converts a synthetic precursor to a noncanonical amino acid (ncAA), which is incorporated into an engineered β-lactamase using an orthogonal translation system (OTS) that suppresses an in-frame stop codon. Active enzymes enable the growth of *E. coli* under selective carbenicillin pressure, while bacteria featuring inactive ones fail to proliferate. (B) Carbamoylase SmLcar catalyzes the hydrolysis of cam-3nY and gives rise to the ncAA L-3-nitro-tyrosine (3nY). (C) Modified selection plasmids with restriction sites allow visual differentiation of SmLcar and SmLcar_GY in a skewed population during serial passaging. (D) Restriction digest followed by agarose gel electrophoresis reveals quantitative patterns for SmLcar and SmLcar_GY during serial passaging. The band at 6085 bp (blue arrow) depicts a linearized plasmid for SmLcar_GY (cut once). The bands at 4016 and 2069 bp (gray arrows) depict two fragments of the SmLcar plasmid (cut twice). The band at 3698 bp (orange arrow) represents the OTS plasmid, which is present in both. (E) Relative base distribution for codons at positions 217 and 329 during the mock selection, obtained from Sanger sequencing pooled plasmids of the mixed populations. Panels A and B were reproduced with permission from ref 16. Copyright 2022, Wiley-VCH GmbH.

small, controlled populations, they can have difficulties in reliably eliciting highly active biocatalysts from gene libraries containing millions of members. For example, genetically modified host cells are prone to escape their confinement by genomic mutations, plasmid rearrangements, or other, unintended mechanisms, thereby increasing false-positive rates during selections.[26,27] Moreover, if products of an enzymatic transformation readily diffuse into the media, growth advantages are not solely confined to library members with high activities but also enable nonbeneficial "hitchhikers" to remain in the population.[28] Thus, to enable a more widespread use of growth-based selections for the directed evolution of biocatalysts, new strategies must prove robust and scalable enough to reliably elicit highly active biocatalysts from large populations.

We have recently showcased a chemical complementation strategy that leverages recoded organisms addicted to non-canonical amino acids (ncAAs, *vide infra*).[16] While our initial work covered rigorous characterization of our strategy's selection principles, its eventual application on a population scale was limited to a small population featuring only 20 different enzyme variants. Here, we demonstrate the robustness and scalability of our growth-based selection system, as it autonomously elicits a diverse panel of efficient biocatalysts from highly diverse populations. Specifically, populations featured up to four simultaneously randomized residues, of which two residues were targeted sequentially before and two positions adjacent to these. Excitingly, serial passaging proved sufficient to amplify inherently scarce variants within these populations, including enzyme variants whose improved performance results from unintended, random mutations that were introduced during PCR cycles or cell division. Attesting to its robustness, we did not observe the enrichment of false positives, as all enzyme variants that bestowed improved

https://doi.org/10.1021/jacsau.4c00070
JACS Au 2024, 4, 1583−1590

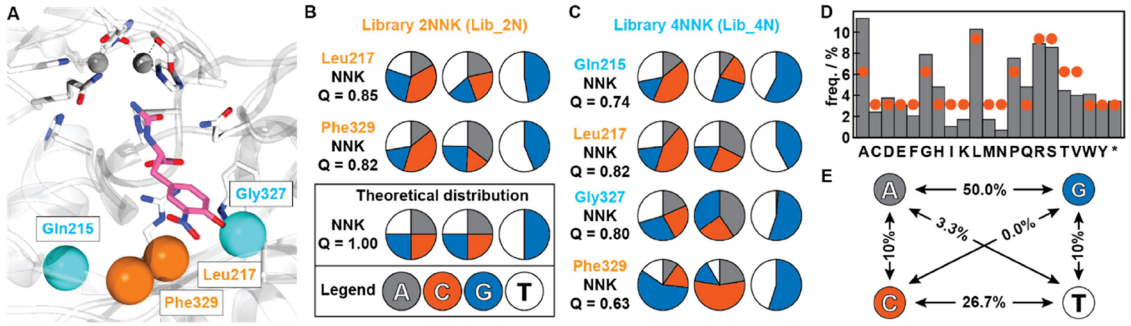


Figure 3. Library design and quality. (A) Close-up view of the active site of SmLcar (PDB: 8APZ). Orange and cyan spheres indicate positions targeted for site-directed mutagenesis. Two bivalent metal ions and residues critical for their binding are given, and the substrate cam-3nY (pink) is docked in the active site. (B, C) Pie charts based on base calls from Sanger sequencing of Lib_2N populations (B) or 71 individual Lib_4N library members (C) depict the relative distribution of bases at each targeted position. *Q*-values are given as a measure of diversity. The theoretical distribution is based on the perfect NNK codon and has a *Q*-value of 1.00. (D) Bar chart depicting the observed distribution of amino acids for Lib_4N (gray bars) and expected theoretical distribution (red dots). (E) Percentage of transitions (left/right) and transversions (up/down, diagonal) observed in Lib_4N based on 30 random mutations found in the sequence data of 71 library members in Lib_4N.

growth properties upon their hosts indeed exhibited greater catalytic performance. Combined, these results not only showcase the impressive potential and simplicity of our growth-based selection system for enzyme engineering but also lay a strong foundation for its application for autonomous exploration of extended evolutionary trajectories of biocatalysts in the future.

RESULTS AND DISCUSSION

Selecting Improved Enzymes from Heavily Skewed Populations

For our recently reported chemical complementation strategy, we first created bacteria, whose lives in the presence of carbenicillin were dependent on the supply and incorporation of a noncanonical amino acid (ncAA) into an engineered β-lactamase. Next, we leveraged the ncAA-dependency of such *recoded bacteria*[29,30] to evolve a biocatalyst that could yield these ncAAs from externally provided, synthetic precursors (Figure 2A).[16] As such, our selections build on the large body of work which employs growth-based selections to engineer aminoacyl-tRNA synthetase (aaRS)/tRNA pairs (=orthogonal translation system).[31] Critically, while both approaches rely on the readthrough of an in-frame stop codon in an antibiotic resistance gene, our platform engineers an enzyme that can provide an ncAA from synthetic precursors rather than an orthogonal translation system. As proof-of-principle, we applied our platform to engineer an L-N-carbamoylase from *Sinorhizobium meliloti* strain CECT 4114, SmLcar[32] to promote the hydrolysis of the *N*-carbamoylated ncAA, cam-L-3-nitro-tyrosine (cam-3nY, Figure 2B). We found that the growth rates of *E. coli* cells in the presence of carbenicillin correlated with the activities of the produced SmLcar variants. Sequentially randomizing positions Leu217 and Phe329 enabled us to identify SmLcar_GY (named after its Leu217Gly and Phe329Tyr substitutions), which displayed shorter lag times under selective conditions and proved >1000 times more proficient than the wild-type carbamoylase.[16] Lastly, we also demonstrated that serial passaging could faithfully elicit SmLcar_GY from a small, controlled population containing 20 enzyme variants.[16]

With the aim of selecting efficient biocatalysts from large libraries, we first set out to gain quantitative insights into the amplification of improved enzyme variants in a heavily skewed population. Toward this end, we constructed selection plasmids, which allow for a straightforward differentiation of SmLcar and SmLcar_GY following restriction digest. In brief, strategically placed restriction sites yield two fragments for SmLcar upon digestion, while only linearizing SmLcar_GY (Figure 2C, see Supporting Information for details). Using these constructs, we performed a mock selection by mixing cultures containing SmLcar and SmLcar_GY in a 100:1 ratio, thus increasing the skewness by an order of magnitude when compared to our previous study. During serial passaging, we grew this skewed population under selective pressure (20 μg/mL carbenicillin) in the presence of the substrate (500 μM cam-3nY), and, after culturing for ∼24 h, performed 100-fold dilutions in fresh selective medium. Restriction digests on isolated plasmids from the starting and selected populations revealed that selection consistently favors the amplification of SmLcar_GY (Figure 2D). In fact, two serial passages sufficed for SmLcar_GY to become dominant. Subjecting the isolated plasmids to Sanger sequencing independently confirmed that the population was essentially devoid of the wild-type carbamoylase following two serial passages (Figure 2E). Combined, these results indicate that the growth advantage more active biocatalysts provide to their hosts suffices to reliably amplify them within a population, even when starting at a considerable disadvantage.

Design, Preparation, and Evaluation of SmLcar Libraries

After the successful enrichment of SmLcar_GY in the mock selection, we set out to retrieve highly active SmLcar variants from increasingly diverse populations. Toward this end, we designed two SmLcar libraries, Lib_2N and Lib_4N, by either randomizing two or four residues that line the binding pocket of SmLcar (Figure 3A). Specifically, Lib_2N targets Leu217 and Phe329 simultaneously, residues which have shown high synergy when *sequentially* randomized in our previous work that elicited SmLcar_GY.[16] Evaluating all 400 possible amino acid combinations *at once* should thus shed further light on potential epistatic interactions between these two residues. Lib_4N additionally includes neighboring residues Gln215 and Gly327, whose amino acid side chains also project toward the binding pocket (Figure 3A). Drastically increasing the genetic diversity from $\approx 10^3$ for Lib_2N to $\approx 10^6$ in Lib_4N should enable us to uncover multiple, new "solutions" toward boosting the activity of SmLcar for the hydrolysis of cam-3nY. Following

https://doi.org/10.1021/jacsau.4c00070
JACS Au 2024, 4, 1583−1590

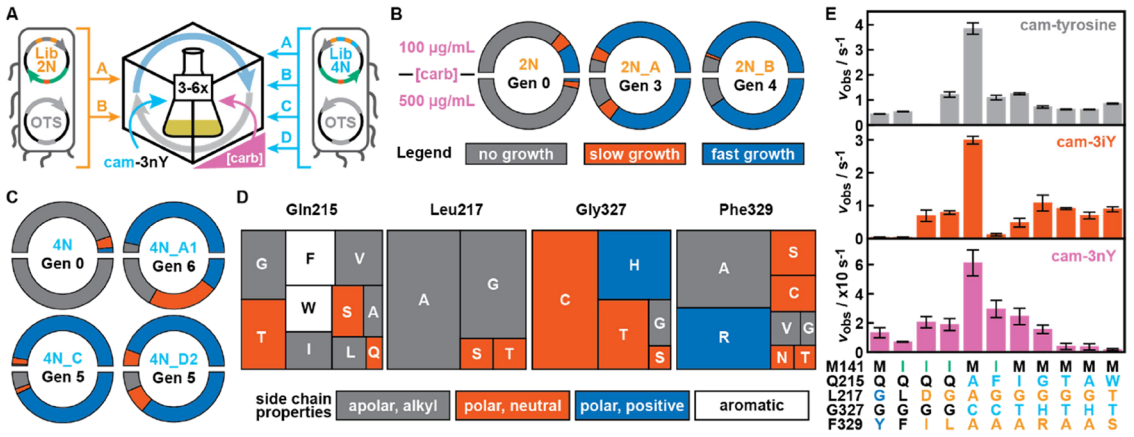


Figure 4. Autonomous selection of proficient SmLcar variants. (A) Workflow for the identification of improved SmLcar variants from highly diverse populations by serial passaging. (B, C) Radial bar charts depicting the distribution in growth phenotypes for the start and end points of 2N populations (B) and 4N populations (C). Distributions follow from growth under 100 (upper half) or 500 (lower half) μg/mL carbenicillin pressure. No growth is defined as failing to exceed an OD_{600} of 0.3 within 48 h; slow growth is defined as exceeding an OD_{600} of 0.3 after 24 to 48 h; fast growth is defined as exceeding an OD_{600} of 0.3 within 24 h. (D) Tree map charts depicting the diversity within amino acids across all four randomized positions in 132 evaluated SmLcar variants (37 unique variants) that bestowed a fast-growing phenotype. (E) Initial rate (v_{obs} in s^{-1}) of reference variant SmLcar_GY and the panel of evolved SmLcar variants for (nc)AA precursors cam-3nY, cam-3iY, and cam-Y. Reactions were performed in triplicates at 25 °C. The corresponding parameters are listed in Table 1.

standard procedures for the preparation of Lib_2N and Lib_4N, we obtained 2.4×10^3 transformants for Lib_2N and 8.4×10^6 transformants for Lib_4N, thus exceeding the theoretical genetic diversity of these libraries by a factor of 2.4 and 8.4, respectively (Table S1).

Following the transformation of libraries into recoded bacteria, we subjected either pooled populations (Lib_2N) or 71 individual library members (Lib_4N) to Sanger sequencing (see Supporting Information for details) to evaluate the diversity of randomized positions and pinpoint potential unplanned mutations introduced during library preparation. Note that the latter makes Sanger sequencing a preferred option over next-generation sequencing (i.e., Illumina), for which random copying errors would have been indiscernible from erroneous base calls.[33] From the obtained sequencing chromatograms, we then calculated the *Q*-values[34] of each NNK stretch, which is a quantitative analysis of library degeneracy, which revealed a high degree of genetic diversity for both libraries (Figures 3B,C and Table S2). Specifically, Lib_2N displays *Q*-values of 0.85 and 0.82, indicating a nearly equal distribution of the desired nucleotides at both NNK stretches. While by comparison Lib_4N scores slightly worse (*Q*-values ranging from 0.63 to 0.82), the genetic diversity across all four targeted positions gives rise to a distribution of amino acids that closely tracks the frequencies expected for NNK codons (Figure 3D). Notably, we were unable to detect thymidine at the second randomized nucleotide in position Gly327 in any of the inspected library members (Figure 3C), reducing the diversity of Lib_4N by 19%.

At the same time, we surmised that relying on three PCRs to amplify (part of) the SmLcar gene for library preparation could have inadvertently increased the diversity of Lib_2N and Lib_4N by accumulating random copying errors. Indeed, when analyzing high-quality Sanger sequencing data of 71 individual Lib_4N variants (Supporting Data File 1), we identified on average approximately 0.7 random mutations per SmLcar gene (0.6 mutations/kb), consistent with the low error rate of the employed high-fidelity DNA polymerase ($\approx 1 \times 10^{-6}$ substitutions per base). The majority (77%) of these mutations were transition substitutions (50% for A ⇄ G and 27% for C ⇄ T), which involve interchanges of purines or pyrimidines (Figure 3E and Table S3). Critically, these random mutations add to the overall genetic diversity of the libraries, with variants featuring beneficial amino acid substitutions having the chance to be amplified over consecutive serial passages. Taken together, the results attest that Lib_2N and Lib_4N are large libraries that display sufficient genetic diversity, which results from randomizing key residues lining the binding pocket of SmLcar and random mutations introduced during library preparation.

Autonomous Selection of Proficient Carbamoylases from Large Libraries

With sufficiently diverse libraries in hand, we next challenged our growth-based selection system to autonomously elicit a panel of highly active SmLcar variants. Following library generation and transformation into recoded hosts, the selection of proficient carbamoylases based on growth advantage is as follows: (1) serial passaging of replicate populations, while increasing the selection pressure from mild to moderate (20, to 50, to 100 μg/mL carbenicillin); (2) plating of enriched variants on LB agar; (3) evaluation of single colonies' growth rate in the plate reader under moderate or high selection pressure (100 or 500 μg/mL carbenicillin), and (4) identifying SmLcar variants that grew faster at the high selection pressure by sequencing (see Supporting Information for details). To probe whether the selection of proficient SmLcar variants is subject to stochasticity (e.g., sampling bias in initial rounds), we performed serial passaging for two Lib_2N populations (2N_A–B) and four Lib_4N populations (4N_A–D) in parallel (Figure 4A). With the exception of 4N_B, which went extinct following the second serial passage, the performed selections thus bestowed on each surviving population a unique *evolutionary history* (Figure S1).

For populations 2N_A and 2N_B, we performed three and four serial passages, respectively, at which point selections had curtailed the genetic diversity of these populations to a few genotypes as judged by Sanger sequencing (Figure S2). As anticipated, the global fitness of the population steadily

https://doi.org/10.1021/jacsau.4c00070
JACS Au 2024, 4, 1583–1590

Table 1. Kinetic Parameters of SmLcar and Evolved Variants for (nc)AA Precursors cam-3nY, cam-3iY, and cam-Y[a]

	cam-L-3-nitro-tyrosine (cam-3nY)		cam-L-3-iodo-tyrosine (cam-3iY)		cam-L-tyrosine (cam-Y)	
enzyme	v_{obs} [s^{-1}]	relative	v_{obs} [s^{-1}]	relative	v_{obs} [s^{-1}]	relative
SmLcar	7.87×10^{-5} (1.66×10^{-5})	1	3.33×10^{-5} (1.10×10^{-6})	1	0.012 (2.59×10^{-4})	1
SmLcar_GY	0.13 (0.034)	1690	0.040 (0.009)	1210	0.44 (0.009)	37
SmLcar_M141I	0.069 (0.002)	880	0.039 (0.0010)	1160	0.52 (0.0088)	44
SmLcar_DI*	0.20 (0.038)	2600	0.69 (0.18)	20,800	0.023 (0.0036)	2
SmLcar_GL*	0.19 (0.042)	2390	0.79 (0.055)	23,700	1.21 (0.11)	103
SmLcar_AACA	0.61 (0.090)	7760	2.99 (0.12)	89,800	3.85 (0.23)	326
SmLcar_FGCA*	0.30 (0.059)	3760	0.12 (0.037)	3500	1.09 (0.094)	93
SmLcar_IGTA	0.24 (0.056)	3100	0.48 (0.13)	14,300	1.25 (0.039)	106
SmLcar_GGHR	0.16 (0.030)	1980	1.08 (0.24)	32,300	0.72 (0.048)	61
SmLcar_TGTA	0.039 (0.021)	500	0.91 (0.026)	27,300	0.63 (0.017)	53
SmLcar_AGHA	0.036 (0.021)	460	0.70 (0.10)	21,000	0.62 (0.012)	53
SmLcar_WTTS	0.017 (0.007)	220	0.89 (0.077)	26,700	0.86 (0.015)	73

[a]Reactions were performed at pH 6 for cam-3nY and pH 8 for cam-3iY and cam-Y. Reactions were performed in triplicate at 25 °C, with a 2 mM substrate and 1 mM $MnCl_2$. Average v_{obs} is given, with the standard deviation in brackets. The asterisk (*) indicates SmLcar variants from 2N and 4N libraries harboring the Met141Ile substitution.

increased with each passage, as judged by the faster growth (=shorter lag times) of aliquots taken after each growth cycle (Figure S2). To quantify this fitness increase, we compared the growth characteristics of 95 individual colonies taken from the start and 95 individual colonies taken from the respective end points of these populations (Figure 4B and Table S4). When subjected to the moderate selection pressure (=100 μg/mL carbenicillin) present at the end of the serial passaging regime, 19% of the colonies taken from the initial populations displayed a fast-growing phenotype—defined as exceeding an OD_{600} of 0.3 within 24 h or less (Figure 4B). This fraction was markedly increased in the selected populations, with 83% and 88% of all single colonies being categorized as fast-growing in 2N_A and 2N_B, respectively. To identify bacteria featuring the most active SmLcar variants, we challenged all variants to grow at a high selection pressure (=500 μg/mL carbenicillin). The difference between the starting and selected populations became more pronounced, as only 2% of the initial library members but 70% and 81% of 2N_A and 2N_B displayed fast growth (Figure 4B).

Following manual inspection of individual growth curves, we selected the 10 best-performing colonies of 2N_A and 2N_B and identified the SmLcar variants they harbored by sequencing. Consistent with our previous evolutionary campaign, 2N_A had converged to SmLcar_GY variants, which all featured the previously identified Leu217Gly and Phe329Tyr substitutions, but differed by (silent) random mutations in their genes (Table S5). Strikingly, sequencing the best-performing carbamoylase variants from 2N_B revealed markedly different genotypes. While SmLcar_GY was only identified once, the other carbamoylase variants featured six times an Asp217/Ile329 combination (SmLcar_DI) and three times Gly217/Leu329 (SmLcar_GL). Upon closer inspection, both SmLcar_DI and SmLcar_GL shared an additional random Met141Ile substitution (Table S5), which was outside of the targeted positions but affected a residue that also lines the binding pocket (Figure S3). Since the otherwise dissimilar variants shared this substitution, we hypothesized that Met141Ile by itself could be a beneficial mutation that provides a growth advantage. To investigate this hypothesis, we prepared SmLcar_M141I and observed that this variant indeed provided a fast-growing phenotype to producing bacteria at 100 μg/mL carbenicillin but a slow-growing phenotype at 500 μg/mL carbenicillin (Figure S3). To confirm whether shorter lag times observed from selected variants translate to higher catalytic activities, we chose SmLcar_DI, SmLcar_GL, and SmLcar_M141I for further kinetic characterization (*vide infra*).

For populations 4N_A, C, and D, we performed five to six serial passages before examining their composition and the growth characteristics of individual colonies. Sanger sequencing of pooled plasmids revealed that selections had curtailed the diversity of initial libraries (Figure S4), with 4N_A and 4N_C showing the apparent enrichment of a handful of phenotypes. Evaluating the growth curves for ∼30 colonies of 4N_A–D attested to the successful selection of fast-growing phenotypes at a moderate selection pressure (100 μg/mL) but also highlighted that 4N_A and 4N_D contained only a small number of variants that could grow rapidly at 500 μg/mL (Table S4). In an attempt to exploit the stochastic nature of selections to improve the fitness of 4N_A and 4N_D, we revived glycerol stocks from their second passages and performed an additional three selective growth−dilution cycles to yield 4N_A2 and 4N_D2 (Figure S1). While these measures did not prove successful for the former, population 4N_D2 had enriched variants that displayed fast growth at the high selection pressure (Table S4).

To compare the apparent fitness increase of all 4N populations, we determined the growth characteristics of 95 individual colonies taken from the start and 349 individual colonies taken from the respective end points, i.e., 4N_A1, 4N_C, and 4N_D2 (Figure 4C and Table S4). Only 3% of colonies taken from the initial populations displayed a fast-growing phenotype when subjected to the moderate selection pressure (100 μg/mL carbenicillin) that is present at the end of the serial passaging regime (Figure 4C). Conversely, our selections successfully enriched fast-growing phenotypes, with 94%, 95%, and 89% of colonies in 4N_A1, 4N_C, and 4N_D2, respectively, exceeding an OD_{600} of 0.3 in under 24 h. Substantially increasing the stringency to 500 μg/mL carbenicillin halted proliferation entirely in the initial population, while 20% of 4N_A1, 83% of 4N_C, and 72% of 4N_D2 colonies still displayed a fast-growing phenotype (Figure 4C). Taken together, these results showcase that serial passages not only effectively discriminate host fitness levels at

https://doi.org/10.1021/jacsau.4c00070
JACS Au 2024, 4, 1583−1590

the applied pressure but also suffice to amplify variants that can thrive under higher selection pressures.

To pinpoint genotypes that were enriched during the selections of 4N libraries, we compared the sequencing data of the SmLcar genes for 73 single colonies prior to serial passages and 132 fast-growing single colonies at the end of each evolution campaign. As expected, the initial population displayed diverse sequences at the randomized positions, with a good representation of each side chain property (Figure S5 and Table S6). Conversely, the diversity at these residues had greatly diminished for the 132 evaluated SmLcar variants that bestowed a fast-growing phenotype following growth−dilution cycles (Figure 4D and Table S7). Combined, the 37 unique sequences obtained offer key insights and commonalities about the type of SmLcar variants that were enriched across these populations (Figure 4D). (1) Gln215 remained the most diverse position by accepting a variety of small, hydrophobic, aromatic, and polar residues, while being devoid of any charged residues. (2) For Leu217, variants featuring small side chains (Ala, Gly) became strongly enriched (89%). This finding is consistent with our previous evolutionary campaign, where we demonstrated that enlarging the binding pocket via a L217G substitution is critical to accommodate cam-3nY in a productive conformation, in which the aromatic side chain is flipped by ∼90°.[16] (3) The substitutions at Gly327 and Phe329 reveal two distinct *solutions*, which are able to drastically boost the growth rates of recoded bacteria (Figure 4D). In one solution, a small hydrophobic or polar residue at position Phe329 is paired with a residue at Gly327 that provides a hydrogen-bond donor at Cβ (i.e., Cys, Ser, or Thr). In the alternative solution, SmLcar variants feature a His residue at position 327, which is almost exclusively combined with a positively charged Arg at position 329.

In Vitro Characterization of Improved SmLcar Variants

Based on the observed genotypes in selected 2N and 4N libraries, we chose a total of 10 SmLcar variants that granted a fast-growing phenotype for further characterization (Tables 1, S5 and S7). In brief, these are SmLcar_M141I as well as SmLcar_DI and SmLcar_GL, which were identified in 2N_B and also harbor this apparently beneficial yet unintended Met141Ile substitution. From the 4N populations, we included the most amplified variant from each surviving population, which are SmLcar_WTTS (4N_A1), SmLcar_IGTA (4N_C), and SmLcar_TGTA (4N_D2). Additionally, we selected representatives for the two solutions identified in the sequencing analysis, namely SmLcar_AACA and SmLcar_GGHR, as well as a hybrid variant, SmLcar_AGHA, which was the only identified variant that features His327 but *not* Arg329. Furthermore, we chose SmLcar_FGCA, whose substitutions not only follow the first motif but also feature the randomly introduced Met141Ile substitution observed in 2N_B variants. Lastly, we included the wild-type and the previously described SmLcar_GY as points of reference for low and high enzyme activities.

Prior to kinetic characterization, we aimed to eliminate the possibility that the fitness increase in hosts harboring evolved SmLcar variants had resulted from undesirable adaptations or escapes. Toward this end, we cloned the SmLcar genes of our panel anew and transformed them into fresh recoded hosts. Manual inspection of individual growth curves showed that even at 500 μg/mL carbenicillin, all evolved variants still presented short lag times of ∼6−12 h, in accordance with the reference variant SmLcar_GY (Figure S6). Critically, the absence of escapes or false positives affirms the robustness of the genotype−phenotype linkage that is central to our growth-based selection platform.

Next, to verify that we have elicited highly active biocatalysts, we purified SmLcar, SmLcar_GY, and all 10 variants of the panel (Figure S7, Table S8) and determined their initial rates to hydrolyze three (nc)AA precursors (2 mM) in vitro (see Table 1 and Supporting Information for details). These are (1) cam-3nY, which was employed throughout the selection campaigns, (2) cam-3-L-iodo-tyrosine (cam-3iY), an alternative *m*-substituted tyrosine analogue, and (3) cam-L-tyrosine (cam-Y). Consistent with previous reports,[16,32] SmLcar displays poor activity for all three substrates, but has a strong preference for cam-Y (v_{obs} = 0.71 min^{-1}) over cam-3nY (0.28 h^{-1}) and cam-3iY (0.12 h^{-1}). Excitingly, all 10 SmLcar variants from our panel outperform the wild-type enzyme for all substrates—typically by 2−4 orders of magnitude depending on the carbamoylated precursor. For example, the most active variant SmLcar_AACA displays initial rates of 0.61 s^{-1} for cam-3nY, 2.99 s^{-1} for cam-3iY, and 3.85 s^{-1} for cam-Y, thus exceeding wild-type activities by a factor of 7760, 89,800, and 326, respectively (Table 1 and Figure 4E). Moreover, selected enzyme variants displayed high activities over the course of the reaction, enabling full conversion of all precursors within 30 min (cam-Y) to ∼2 h (cam-3nY) at 0.05 mol % catalyst loading (Figures S8−9).

A comparison of the results obtained across all 10 SmLcar variants and three substrates provides some valuable insights (Table 1 and Figure 4E). (1) The selected carbamoylases are highly efficient catalysts, which outperform the previously evolved SmLcar_GY for at least two of the tested precursors. (2) We do not observe the selection of *specialists* that display high activities only for cam-3nY, the precursor used during serial passaging. In fact, SmLcar variants featuring substitutions TGTA, AGHA, and WTTS show significantly higher activities for cam-3iY and cam-Y than cam-3nY. (3) Conversely, we observe the selection of a number of *generalists*, which outperform SmLcar_GY for all three substrates. Specifically, SmLcar_GL, AACA, IGTA, and GGHR display comparable efficiencies on all substrates, suggesting that these substitutions present general catalytic solutions for the conversion of carbamoylated tyrosine and *m*-substituted analogues. (4) Lastly, the randomly introduced Met141Ile substitution is highly beneficial, with additional substitutions at the targeted positions in the 2N and 4N libraries further boosting its activity for cam-3nY. Variants featuring the Met141Ile substitution also show diverse reactivity profiles: SmLcar_FGCA maintains the reactivity profile of the random Met141Ile substitution, SmLcar_DI selectively converts *m*-substituted tyrosine analogues, and SmLcar_GL represents a generalist enzyme.

CONCLUSIONS

Strategies to identify improved enzyme variants from large libraries are highly sought-after in the directed evolution of (synthetically useful) biocatalysts. Here, we demonstrate the robustness and scalability of a previously described growth-based selection platform by autonomously eliciting a diverse panel of efficient biocatalysts from large and diverse populations. Specifically, we used the chemical complementation of recoded hosts to investigate epistatic interactions of two or four simultaneously targeted positions within a carbamoy-

https://doi.org/10.1021/jacsau.4c00070
JACS Au 2024, 4, 1583−1590

lase. Despite starting from replicate populations, we observed the enrichment of carbamoylase variants with distinct substitutions—some of which were the result of random copying errors. Strikingly, each characterized carbamoylase outperformed the wild-type enzyme by at least 2 orders of magnitude, thus presenting a different "solution" to the same enzymatic "problem".

In its current setup, our growth-based selection platform is attractive for its application in enzyme engineering campaigns. First, our platform is fast, easy to implement, and requires no specialized and/or expensive equipment. Selections encompass serial passaging for ∼1 week under increasingly stringent conditions, with carbenicillin serving as a tunable selection pressure and cell survival as a common readout. Second, the platform is robust and high-throughput. We evolved highly efficient biocatalysts without encountering hitchhikers or cells that had escaped their confinement. With the largest library in this work containing $>10^6$ enzyme variants, its scalability appears only limited by transformation efficiency. Third, rather than enzymes specialized for the substrate used in selections, the platform elicits overall catalytic improvements. In fact, we retrieved a number of generalist enzymes that showed drastically increased rates on never-before-seen substrates. Fourth, our growth-based selection strategy is flexible in nature. Centered around the enzyme-catalyzed production of ncAAs from synthetic precursors, the platform should be applicable to engineering mechanistically diverse enzyme classes that can yield tyrosine- or phenylalanine-based ncAAs (Figure S10). To allow for an even broader application, we envision the construction of recoded organisms,[29,35] whose life has been made dependent on the presence of structurally diverse ncAAs that have previously been incorporated into proteins in vivo.[36]

Lastly, we anticipate that interfacing selections based on complementing recoded bacteria with gene-directed hypermutation strategies will result in the construction of a versatile continuous evolution platform.[37,38] In the future, such a system should enable us to navigate the sequence space of diverse biocatalysts autonomously and uncover highly proficient enzymes featuring exceedingly unpredictable mutations, which are far beyond our current rationalization.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacsau.4c00070.

Supporting data, experimental procedures and methods, and sequences of primers and proteins (PDF)

Random mutation analysis (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Clemens Mayer − *Stratingh Institute for Chemistry, University of Groningen, 9747 AG Groningen, The Netherlands;* orcid.org/0000-0002-6495-9873; Email: c.mayer@rug.nl

Author

Suzanne C. Jansen − *Stratingh Institute for Chemistry, University of Groningen, 9747 AG Groningen, The Netherlands;* orcid.org/0000-0002-2507-9072

Complete contact information is available at: https://pubs.acs.org/10.1021/jacsau.4c00070

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.M. and S.C.J. are thankful to R. Rubini for the guidance with the chemical complementation workflow and to T.R. Oppewal for the help with library preparation. C.M. acknowledges the NWO (ENW-M grant OCENW.M20.278 and ENW-XS grant 01259958) for funding.

■ REFERENCES

(1) Bornscheuer, U. T.; et al. Engineering the third wave of biocatalysis. *Nature* **2012**, *485*, 185−194.

(2) Sheldon, R. A.; Woodley, J. M. Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **2018**, *118*, 801−838.

(3) Turner, N. J. Directed evolution drives the next generation of biocatalysts. *Nat. Chem. Biol.* **2009**, *5*, 567−573.

(4) Arnold, F. H. Directed Evolution: Bringing New Chemistry to Life. *Angew. Chem., Int. Ed.* **2018**, *57*, 4143−4148.

(5) Packer, M. S.; Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Gen.* **2015**, *16*, 379−394.

(6) Zeymer, C.; Hilvert, D. Directed Evolution of Protein Catalysts. *Annu. Rev. Biochem.* **2018**, *87*, 131−157.

(7) Wang, Y.; et al. Directed Evolution: Methodologies and Applications. *Chem. Rev.* **2021**, *121*, 12384−12444.

(8) Xiao, H.; Bao, Z.; Zhao, H. High Throughput Screening and Selection Methods for Directed Enzyme Evolution. *Ind. Eng. Chem. Res.* **2015**, *54*, 4011−4020.

(9) McLure, R. J.; Radford, S. E.; Brockwell, D. J. High-throughput directed evolution: a golden era for protein science. *Trends Chem.* **2022**, *4*, 378−391.

(10) Taylor, S. V.; Kast, P.; Hilvert, D. Investigating and Engineering Enzymes by Genetic Selection. *Angew. Chem., Int. Ed.* **2001**, *40*, 3310−3335.

(11) Acevedo-Rocha, C. G.; Agudo, R.; Reetz, M. T. Directed evolution of stereoselective enzymes based on genetic selection as opposed to screening systems. *J. Biotechnol.* **2014**, *191*, 3−10.

(12) Li, Z.; Deng, Y.; Yang, G.-Y. Growth-coupled high throughput selection for directed enzyme evolution. *Biotechnol. Adv.* **2023**, *68*, No. 108238.

(13) Rix, G.; et al. Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nat. Commun.* **2020**, *11*, 5644.

(14) Wu, S.; et al. A growth selection system for the directed evolution of amine-forming or converting enzymes. *Nat. Commun.* **2022**, *13*, 7458.

(15) Dulchavsky, M.; et al. Directed evolution unlocks oxygen reactivity for a nicotine-degrading flavoenzyme. *Nat. Chem. Biol.* **2023**, *19*, 1406−1414.

(16) Rubini, R.; Jansen, S. C.; Beekhuis, H.; Rozeboom, H. J.; Mayer, C. Selecting Better Biocatalysts by Complementing Recoded Bacteria. *Angew. Chem., Int. Ed.* **2023**, *62*, No. e202213942.

(17) Russ, W. P.; et al. An evolution-based model for designing chorismate mutase enzymes. *Science* **2020**, *369*, 440−445.

(18) Song, W. J.; Tezcan, F. A. A designed supramolecular protein assembly with in vivo enzymatic activity. *Science* **2014**, *346*, 1525−1528.

(19) Fahrig-Kamarauskait, J.; et al. Evolving the naturally compromised chorismate mutase from Mycobacterium tuberculosis to top performance. *J. Biol. Chem.* **2020**, *295*, 17514−17534.

(20) Yano, T.; Oue, S.; Kagamiyama, H. Directed evolution of an aspartate aminotransferase with new substrate specificities. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 5511−5515.

(21) Chen, J.; Wang, Y.; Zheng, P.; Sun, J. Engineering synthetic auxotrophs for growth-coupled directed protein evolution. *Trends Biotechnol.* **2022**, *40*, 773−776.
(22) Michener, J. K.; Smolke, C. D. High-throughput enzyme evolution in Saccharomyces cerevisiae using a synthetic RNA switch. *Metab. Eng.* **2012**, *14*, 306−316.
(23) Hallberg, Z. F.; Su, Y.; Kitto, R. Z.; Hammond, M. C. Engineering and In Vivo Applications of Riboswitches. *Annu. Rev. Biochem.* **2017**, *86*, 515−539.
(24) Tay, S. B.; Ang, E. L. in *Protein Eng*, 2021, 29−55.
(25) Miller, C. A.; Ho, J. M. L.; Bennett, M. R. Strategies for Improving Small-Molecule Biosensors in Bacteria. *Biosensors* **2022**, *12*, 64.
(26) Lee, J. W.; Chan, C. T. Y.; Slomovic, S.; Collins, J. J. Next-generation biocontainment systems for engineered organisms. *Nat. Chem. Biol.* **2018**, *14*, 530−537.
(27) Roth, T. B.; Woolston, B. M.; Stephanopoulos, G.; Liu, D. R. Phage-Assisted Evolution of Bacillus methanolicus Methanol Dehydrogenase 2. *ACS Synth. Biol.* **2019**, *8*, 796−806.
(28) Jones, K. A.; Snodgrass, H. M.; Belsare, K.; Dickinson, B. C.; Lewis, J. C. Phage-Assisted Continuous Evolution and Selection of Enzymes for Chemical Synthesis. *ACS Centr. Sci.* **2021**, *7*, 1581−1590.
(29) Tack, D. S.; et al. Addicting diverse bacteria to a noncanonical amino acid. *Nat. Chem. Biol.* **2016**, *12*, 138−140.
(30) Rubini, R.; Mayer, C. Addicting Escherichia coli to New-to-Nature Reactions. *ACS Chem. Biol.* **2020**, *15*, 3093−3098.
(31) Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **2017**, *550*, 53−60.
(32) Martínez-Rodríguez, S.; Clemente-Jiménez, J. M.; Rodríguez-Vico, F.; Las Heras-Vázquez, F. J. Molecular Cloning and Biochemical Characterization of L -N-Carbamoylase from Sinorhizobium meliloti CECT4114. *Microb. Physiol.* **2005**, *9*, 16−25.
(33) Pfeiffer, F.; et al. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci. Rep.* **2018**, *8*, 10950.
(34) Sullivan, B.; Walton, A. Z.; Stewart, J. D. Library construction and evaluation for site saturation mutagenesis. *Enzyme Microb. Technol.* **2013**, *53*, 70−77.
(35) Koh, M.; Yao, A.; Gleason, P. R.; Mills, J. H.; Schultz, P. G. A General Strategy for Engineering Noncanonical Amino Acid Dependent Bacterial Growth. *J. Am. Chem. Soc.* **2019**, *141*, 16213−16216.
(36) Dumas, A.; Lercher, L.; Spicer, C. D.; Davis, B. G. Designing logical codon reassignment - Expanding the chemistry in biology. *Chem. Sci.* **2015**, *6*, 50−69.
(37) Morrison, M. S.; Podracky, C. J.; Liu, D. R. The developing toolkit of continuous directed evolution. *Nat. Chem. Biol.* **2020**, *16*, 610−619.
(38) Molina, R. S.; et al. In vivo hypermutation and continuous evolution. *Nat. Rev. Methods Primers* **2022**, *2*, 37.

https://doi.org/10.1021/jacsau.4c00070
JACS Au 2024, 4, 1583−1590